International Journal of Civil, Structural, Environmental and Infrastructure Engineering Research and Development (IJCSEIERD)
ISSN(P): 2249-6866; ISSN(E): 2249-7978
Vol. 7, Issue 1, Feb 2017, 27-36

PRELIMINARY AND FAST RISK ASSESSMENT OF DEBRIS-FLOW HAZARDS USING DIFFERENT ARTIFICIAL NEURAL NETWORKS

LI-JENG HUANG[1] & TZY-WEI CHEN[2]

[1]Associate Professor, Department of Civil Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan, R.O.C

[2] Master Student, Department. of Civil Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan, R. O. C

ABSTRACT

Researches revealed that prediction of occurrence of debris-flow hazards can be achieved by artificial intelligence techniques. In the paper three different artificial neural networks (ANNs), i.e., back-propagation neural network (BPNN), radial basis function neural network (RBFNN) and probabilistic neural network (PNN), are employed for risk assessment of debris-flow hazards. Case of watershed area of Chen-Yu-Lan River in Nantou, Taiwan, studied by Hsiao (2003) who employed an integrated technique of digital terrain model (DTM), image processing schemes, and geographical information system (GIS) is considered. ANN models with 15 influence factors and with 7 influence factors obtained by principal component analysis (PCA) are tested, respectively. In the BPNN topology four different learning rules of MATLAB nntool are selected. It is shown that the percentage of accuracy of debris-flow hazard prediction can be up to 90.48% using BPNN models with totally 15 influence factors. Present study revealed that integration of DTM, GIS and ANN techniques can construct a preliminary and fast risk asse assessment of debris flow hazards for watershed areas.

KEYWORD: Artificial Neural Networks, Debris-flow, Multi-criteria, Risk Assessment & Hazard Prediction

Received: Dec 01, 2016; **Accepted:** Dec 31, 2016; **Published:** Jan 07, 2017; **Paper Id.:** IJCSEIERDFEB20172

Original Article

1. INTRODUCTION

Prediction and prevention of debris-flow hazards have become an important issue for those countries with a lot of mountains and rainfalls such as Japan, USA and Taiwan, etc (Jan & Shen, 1993; Wu et al., 2006; Lin, 2006). Some engineering constructions can be built for prevention of the stony and muddy flows, such as slit and comb dams; while some precaution and warning systems are also useful for mitigation of disasters in time (Jan, 2000; Huang, 2001). From the points of view of disaster prevention a well-established quick and accurate risk assessment system for evaluating the possibility or potential of occurrence of debris-flow hazards will reduce a lot of loss of human being and economy in a region.

However, disasters caused from debris-flows are affected by compound factors. For example, local topographic, meteorological, geological and hydrologic conditions all play important role on the occurrence (Takahashi, 1991). From the previous studies it is well known that accurate analyses of the watershed area conditions (Yu, et al., 1991; Jan, 2000) offer preliminary risk assessment of debris-flow hazards occurring in that watershed area. However, there are also a lot of influence factors of a river in a watershed area that should be employed for risk assessment of debris-flows. A lot of techniques based on theory of probability, geography

information system (GIS), etc. have been developed and reviewed (Huang 2015).

It is interesting that some researchers started to apply techniques related to artificial intelligence (AI) to debris-flow hazard assessment and prediction: expert systems (Lin, 2000); artificial neural networks (Hsiao, 2003), fuzzy reasoning system (Chen, 2006); case-based system (Tsai, 2007) and grey system theory (Lin, 2016).

Artificial neural networks (ANNs) have been widely developed and applied to classification and regression of data analysis (Hagan, et al.1996; Hecht-Nielsen, R. 1990; Jang, et al. 1997). In the machine learning sciences ANNs are recognized as supervised process in which the networks are trained and adjusted using part of data and learning rule until the error between target and simulated output becomes small to an extent(Harrington, 2012). Then a set of testing data can be employed to investigate the validity of prediction using this trained network. In the application of debris-flow hazards the ANNs can be employed as classifiers, i.e., prediction of occurrence or not.

In the paper three different artificial neural networks (ANNs), i.e., back-propagation neural network (BPNN), radial basis function neural network (RBFNN) and probabilistic neural network (PNN), are employed for risk assessment of debris-flow hazards of watershed area of Chen-Yu-Lan River in Nantou, Taiwan. The original analysis data base is obtained from Hsiao (2003) who employed an integrated technique of digital terrain model (DTM), image processing schemes, and geographical information system (GIS). At first ANN models with 15 influence factors are employed and then ANN models with 7 influence factors obtained by principal component analysis (PCA) are tested. In the BPNN topology four different learning rules such as *traingd, traingdm, traingda* and *trainlm* of MATLAB toolbox *nntool* were selected.

2. RESEARCH METHODS

2.1 Geographical Information System

A geographical information system (GIS) integrates the functions of image processing, computer-aided graphics, remote sensing techniques, data base management and data analysis of geographical information. The GIS can be employed to preliminary and fast risk assessment of debris-flow hazards and the advantages come from its global investigation and huge information. In this research the analysis data is obtained from Shiao (2003) which was analyzed using Mapinfo 6.0.

2.2 Artificial Neural Networks

The basic idea of application of artificial neural networks (ANNs) to risk assessment of debris-flow hazards can be explained as building a well-trained computational model (including topology and learning rules) to calculate the possibility of occurrence of debris-flow disaster from some pre-selected influence factors. Using lots of known inputs and outputs we can train and test the computational model (adjusting the weights in neural networks) using parts of sample data. Risk of debris-flow hazard of a new area can be assessed if the computational model has been well-trained from known data of inputs and outputs.

In this research we employed three kinds of supervised artificial neural networks as follows:

- **Back-Propagation Neural Networks (BPNN):**

$$H_j = \sum_{i=1}^{NI} f(w_{ji}X_i) + b_j, \qquad i = 1,2,\cdots NI, \quad j = 1,2,\cdots NH \tag{1}$$

$$Y_k = \sum_{j=1}^{NH} f(w_{kj}H_j) + b_k, \qquad k = 1,2,\cdots NO, \quad j = 1,2,\cdots NH \tag{2}$$

The activating function using in the BPNN is the Sigmoid function:

$$f(x) = \frac{1}{1+e^{-x}} \tag{3}$$

The root mean squared error (RMSE) of training samples can be calculated as

$$\min E = \sum_{p=1}^{NP} E^p = \sum_{p=1}^{NP} \frac{1}{2}\left\|\tilde{y}_k^p - \tilde{d}_k^p\right\| \quad (p = 1,2,\cdots,NP; \quad k = 1,2,\cdots,NO) \tag{4}$$

- ***Generalized Steepest Decent (GD):***

The learning rule can be written as

$$\Delta w_{ij}(t+1) = w_{ij}(t+1) - w_{ij}(t) = -\eta \frac{\partial E}{\partial w_{ij}}(t) + \mu\, \Delta w_{ij}(t) \tag{5}$$

where η deotes the learning rate and μ denotes the momentum factor.

- ***Generalized Steepest Decent Including Momentum (GDM):***

The learning rule can be written as

$$\Delta w_{ij}(t+1) = w_{ij}(t+1) - w_{ij}(t) = -\alpha \frac{\partial E}{\partial w_{ij}}(t) + \mu \Delta w_{ij}(t) \tag{6}$$

where α deotes the learning rate and μ denotes the momentum factor.

- ***Generalized Steepest Decent with Adjustable Learning Rate (GDA):***

In this algorithm the basic learning rule is the same as Eq. (6) but adding a conditional judgment. When stable learning can be kept under a learning rate then the learning rate is increased, otherwise it is decreased. The learning rate increment and decrement can be denoted as ζ_{inc} and ζ_{dec} (Hagen et al, 1996).

- ***Levenberg-Marquardt (LM):***

The learning rule can be written as

$$\Delta w_{ij}(t+1) = w_{ij}(t+1) - w_{ij}(t) = -[[J]^T[J] + \lambda[I]]^{-1}[J]^T\{e\} \tag{7}$$

where λ denotes a constant to assure the inversion of matrix, and the learning rule becomes Gauss-Newton algorithm whenλ= 0, while it approaches GD with small learning rate with large λ. (Hagen et al, 1996).

- **Radial Basis Function Neural Networks (RBFNN):**

In RBFNN algorithm, the relations from inputs to hidden neurons and the relation from hidden layer to outputs can be expressed as

$$H_j = \exp(-\gamma\left\|X_i - C_j\right\|), \qquad i = 1,2,\cdots NI, \quad j = 1,2,\cdots NH \tag{8}$$

$$Y_k = \sum_{j=0}^{NH} w_{jk} H_j(X_i), \qquad k = 1,2,\cdots NO, \quad j = 1,2,\cdots NH \tag{9}$$

where γ denotes the spread constant of Gau Gaussian functions.

- **Probabilistic Neural Networks (PNN):**

  The probability density function is mployed in Bayesian classifier as

$$P_j(X_i) = \exp[\frac{-\sum (X_i - W^{ij})^2}{2\,\sigma^2}], \quad i = 1,2,\cdots, NI, \quad j = 1,2,\cdots, NP \tag{10}$$

  The output is calculated by method of centroid:

$$Y_k = \frac{\sum_{j=1}^{NJ} Z_j y_j P_j(X)}{\sum_{j=1}^{NJ} Z_j P_j(X)}, \qquad k = 1,2,\cdots NO, \quad j = 1,2,\cdots NP \tag{11}$$

2.3 Principal Component Analysis (PCA):

A famous technique in multivariate statistical analysis for reducing dimension of variables is the principal component analysis (PCA) (Abdi & Williams, 2010). Assuming {X} denotes the original input vector of $N \times 1$, then the new input vector of reduced dimension $L \times 1$ can be obtained by

$$\{\tilde{X}\} = [W_L]\{X\} = [\Sigma_L]\{V\} \tag{12}$$

where $[\Sigma_L] = [I]_{L\times N}[\Sigma]$ and $[W_L]$ is a sub-matrix of orthogonal matrix $[W]$. The reduced dimensional system can be built by keeping some larger eigen-values and deleting those corresponding to smaller eigen-values.

3. CASE STUDY OF RISK ASSESSMENT OF DEBRIS-FLOW HAZARDS

3.1 Research Scope

Prediction of debris-flow hazards of watershed area of Chen-Yu-Lan River in Nantou County, Taiwan is considered. The original analysis data base is obtained from Hsiao (2003) who employed an integrated technique of digital terrain model (DTM), image processing schemes, and geographical information system (GIS). Different ANNs are employed for preliminary and fast risk assessment.

3.2 Influence factors

Fifteen influence factors are involved in the assessment of debris-flow hazard of the Chen-Yu-Lan River (Hsiao, 2003).

- **Watershed Conditions**

  A1. Watershed area (WA)

  A2. Gradient in X-direction of watershed (GXW)

  A3. Gradient in Y-direction ofwatershed (GYW)

A4. Length of main stream (LS)

- **Deposit Conditions**

B1. Volumetric measure of watershed (VMW)

B2. Maximal slope of watershed (MSW)

B3. Averaged slope of watershed (ASW)

- **Topographic Conditions**

C1. Form factor of watershed (FFW)

C2. Circularity ratio of watershed (CRW)

C3. Elongation ratio of watershed (ERW)

C4. Relative height of watershed (RHW)

C5. Coefficient of variation of height (CVH)

C6. Relative height of main stream (RHS)

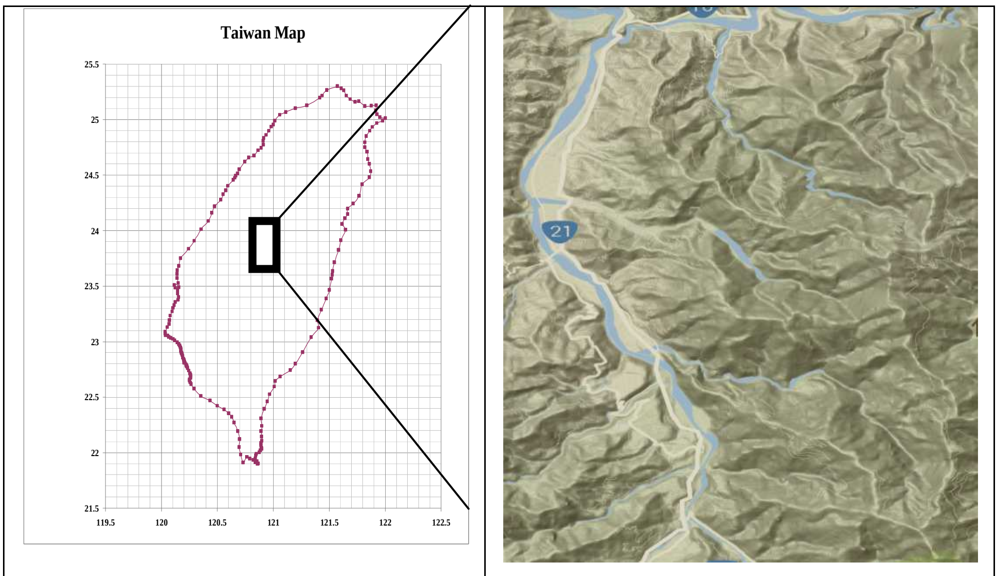
C7. Curvature of main stream (CS)

C8. Averaged slope of main stream (ASS)

4. ANALYSIS RESULTS

4.1 Description of Watershed Area of Chen-Yu-Lan River

Chen-Yu-Lan River is a river in Nantou County, Taiwan. The watershed area is about 450 m^2, and its length is 42.4 *km*, and the average slope is 5%. (Figure 1).



(a) Location of Watershed Area of Chen-Yu-Lan River in Taiwan Map | **(b) Topological Map of Area of Chen-Yu-Lan River (Google Map)**

Figure 1: Locations of Watershed Area of Chen-Yu-Lan River in Nantou County, Taiwan

4.2 Original Data Employed for ANN Analysis

The original data for the current risk assessment of debris-flow hazards in watershed of Chen-Yu-Lan River is obtained from the Appendix B in reference of Hsiao (2003). The data were built based on three basic data bases: (1) Digital Terrain Model (with resolution $40m\times 40m$); (2) River system data; and (3) Data of rivers related to debris-flows caused by TORAJI Typhoon, 2001. The data were achieved using the following criteria: (1) The smallest rainfall thresh-hold is 100; (2) the watershed area is limited within 0.15 to 4.5 m^2; (3) the length of main stream is greater than 200 m. The maximal and minimal values of 15 influence factors are shown in Table 1.

Table 1: Data Range of 15 Influence Factors for Risk Assessment of Debris-Flow Hazards in Watershed Area of Chen-Yu-Lan River

	A1	A2	A3	A4	B1	B2	B3	C1	C2	C3	C4	C5	C6	C7	C8
	WA	GXW	GYW	LS	VMW	MSW	ASW	FFW	CRW	ERW	RHW	CVH	RHS	CS	ASS
	m^2			km		Deg.	Deg.				m		m		deg.
Max	4.5	1	1	5.2	1	89	89	1	1	1	2	0.5	1.5	2	89
Min	0.15	-1	-1	0.2	0	0	0	0	0	0	0	0	0	1	0

In order to obtain better output results in ANNs, the original data should be normalized as

$$X_{new} = \frac{X_{old} - X_{\min}}{X_{\max} - X_{\min}} \tag{13}$$

To transform the original data to new data within the region [0, 1]. The normalized data in the table of Hsiao (2003) includes 389 samples and 15 influence factors. Only 245 samples are those with assured occurrence or no occurrence of hazards. In the 245 samples we selected 71 samples of occurrence and randomly selected 71 samples of no occurrence to build up totally 142 basic data for ANN study.

In the following ANN simulations we randomly selected 100 samples for training set and the remaining 42 samples for testing. Both the training set and the testing set involve the cases of occurrence (marked by 1) and no occurrence of debris-flow (marked by 0).

Table 2: Normalized Data of 245 Identified Samples with 15 Influence Factors (Hsiao, 2003)

	A1	A2	A3	A4	B1	B2	B3	C1	C2	C3	C4	C5	C6	C7	C8	
	WA	GXW	GYW	LS	VMW	MSW	ASW	FFW	CRW	ERW	RHW	CVH	RHS	CS	ASS	
1	0.035	0.897	0.685	0.022	0.603	0.612	0.491	0.232	0.409	0.543	0.394	0.139	0.120	0.176	0.378	0
2	0.060	0.653	0.100	0.102	0.514	0.695	0.492	0.201	0.386	0.506	0.578	0.153	0.348	0.096	0.418	0
30	0.361	0.500	0.807	0.583	0.464	0.536	0.352	0.169	0.315	0.464	0.603	0.327	0.551	0.277	0.198	1
244	0.312	0.771	0.262	0.317	0.539	0.619	0.454	0.244	0.497	0.558	0.824	0.311	0.710	0.112	0.358	0
245	0.143	0.767	0.340	0.098	0.460	0.643	0.493	0.480	0.584	0.781	0.538	0.140	0.306	0.194	0.415	0

4.3 ANNs Results Using 15 Influence Factors

We first conducted risk assessment of debris-flow hazards using various ANNs considering 15 influence factors. All the analyses were executed using MATLAB toolbox *nntool*. The topology of BPNN is NI-NH-NO = 15-8-1, where the number of neurons in hidden layer is calculated from NH = (NI+NO)/2. The variation of RSME with epoch of training is shown in Figure2 where the results of four different learning rules are employed. As expected the Levenberg-Marquardt

algorithm gives the quickest training. The predicted results and percentage of accuracy obtained from four BPNNs, two RBNNs and PNN are shown in Table 3 along with those obtained by Hsiao (2003) using BPNN with 15-12-20-1. From our analysis BPNNs using only single hidden layer and 8 neurons can achieve percentage of accuracy up to 90.48% nearly the same or better than those using two hidden layers of Hsiao (2003). In the cases the best predicted results coming from BPNN-2(GDM), BPNN-3 (GDA) and RBFNN-2.

Table 3: Predicted Results using Different ANNs and Learning Rules (15 Influence Factors)

	Hsiao (2003)	BPNN-1 (GD)	BPNN-2 (GDM)	BPNN-3 (GDA)	BPNN-4 (LM)	RBFNN-1	RBFNN-2	PNN
No. of Inputs	15	15	15	15	15	15	15	15
Network	15-12-20-1	15-8-1	15-8-1	15-8-1	15-8-1	SC=1000	SC=10000	
No. Errors		5	4	4	5	6	4	6
Incorrectly Predicted Testing Samples		4 11 23 28 41	15 23 28 41	11 23 28 41	11 23 28 38 41	4 11 23 27 28 41	4 23 28 41	4 11 23 28 33 41
Percentage of Accuracy	88.10%	88.10%	90.48%	90.48%	88.10%	85.71%	90.48%	85.71%
RMSE of Testing		0.042	0.042	0.042	0.049	0.048	0.046	0.058

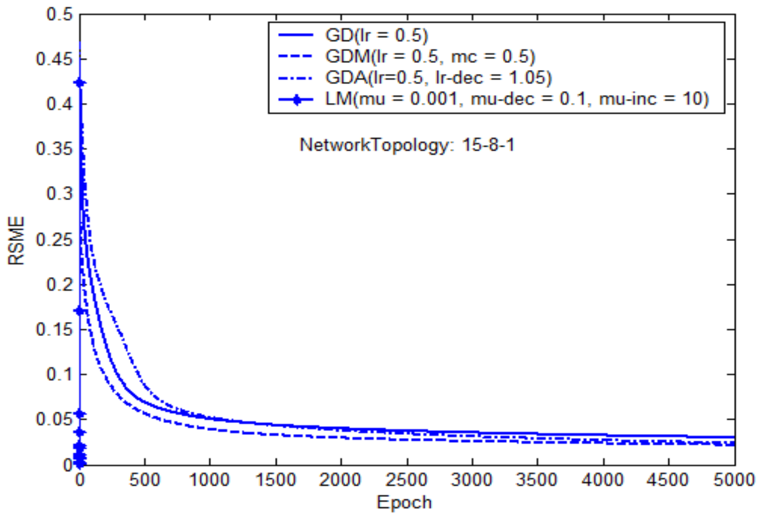


Figure 2: The Variation of RSME with Epoch of Training for Four Learning Rules of BPNN Using 15 Influence Factors

4.4 PCA-ANN Results Using 7 Influence Factors

We then employ PCA of MATLAB to reduce the dimension of input vector. Considering keeping those inputs with minimal variation smaller than 3% (min_frac = 0.03) we found that NI has been reduced from 15 to 7. In order to compare with the results of section 4.3 employ the BPNN topology 7-8-1 with the same 4 learning rules. The variation of RSME with epoch of training is shown in Figure3. The Levenberg-Marquardt algorithm again gives the quickest training. The predicted results and percentage of accuracy obtained from four BPNNs, two RBNNs and PNN are shown in Table 4 along with those obtained by Hsiao (2003) using BPNN with 6-2-2-1. In this case our the best prediction of our analysis of ANNs is only 69.04%. This is lower than 88.10% by Hsiao (2003) using 6 input vectors (A3, A4, B3, C5, C7 and C8)

obtained from multivariate statistical analysis and using software PCNeuron, and 85% by Chen (2016) using 7 input variables (A1, A4, B2, B3, C4, C6, and C8) obtained by inspection using MATLAB and VBA. The reason that using PCA for reduction of input variable leads to poorer prediction maybe come from the possible interference between influence factors of watershed, deposit and topographical conditions. However, nearly 70% accuracy can be achieved using only 7 input variables by ANNs (e.g. BPNN-2 and PNN). Further study can be attempted to investigate the systematic approach for reduction of the dimension of input vector of ANNs.

Table 4: Predicted Results Using Different ANNs and Learning Rules (7 Influence Factors Obtained from PCA with 0.03)

	Hsiao (2003)	BPNN-1 (GD)	BPNN-2 (GDM)	BPNN-3 (GDA)	BPNN-4 (LM)	RBFNN-1	RBFNN-2	PNN
No. of Inputs	6	7	7	7	7	7	7	7
Network	6-2-2-1	7-8-1	7-8-1	7-8-1	7-8-1	SC=1000	SC=10000	
No. Errors		16	13	16	14	19	15	13
Incorrectly Predicted Testing Samples		3 6 9 12 13 16 18 28 29 30 31 32 33 35 38 40	6 9 12 16 18 20 23 28 30 31 32 33 38	3 6 9 12 18 20 23 28 29 30 31 32 33 35 38 40	9 12 13 16 18 23 28 29 30 31 32 33 35 38	2 3 6 9 12 14 17 19 21 25 28 30 31 32 33 38 39 40 42	3 6 9 11 12 20 23 28 29 30 31 32 33 35 38	3 6 9 12 18 28 29 30 31 32 33 38 40
Percentage of Accuracy	88.10%	61.90%	69.05%	61.90%	66.67	54.76%	64.29%	69.05%
RMSE of Testing		0.087	0.075	0.083	0.085	0.102	0.078	0.086

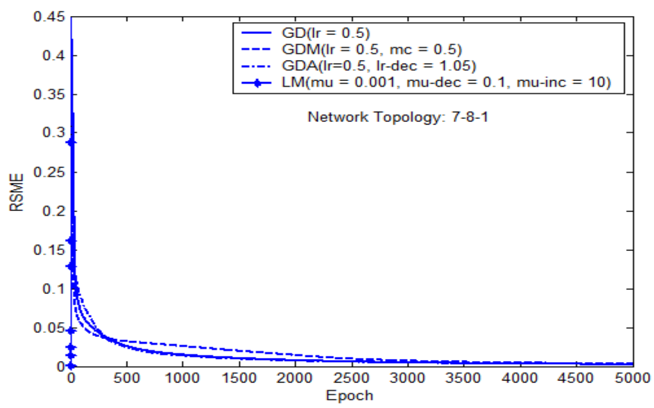


Figure 3: The Variation of RSME with Epoch of Training for Four Learning Rules of BPNN using 7 Influence Factors

5. CONCLUSIONS

Three different artificial neural networks (ANNs), i.e., back-propagation neural network (BPNN), radial basis function neural network (RBFNN) and probabilistic neural network (PNN), are employed for preliminary and fast risk assessment of debris-flow hazards of watershed area of Chen-Yu-Lan River in Nantou, Taiwan. Results depicted that:

- ANN models with 15 influence factors can provide up to 90.48% percentage of accuracy in prediction of debris-flow hazards;
- ANN models with 7 influence factors obtained by principal component analysis (PCA) give percentage of accuracy in prediction only lower than 70%.
- The predicted values obtained by RBFNNs and PNN are more stable than BPNNs but might be with lower percentage of accuracy of prediction.
- Integration of DTM, GIS and ANNs can offer a preliminary and fast risk assessment of debris-flow hazards in a watershed area. It is suggested that engineers and managers can employ various ANNs to obtain comparative and confident predicted results.

REFERENCES

1. *Takahashi, T. (1991). Debris Flow, Balkema, Rotterdam.*
2. *Jan, C. D. and Shen, H. W. (1993). A review of debris flow analysis, Proc. XXV Congress, IAHR, 3, 25-32.*
3. *Wu, J. M., Yu, F. C. and Chao, C. M. (2006). The earth disasters of slope lands in Taiwan, J. of Envir. and Manag., VII(1), 86-98..*
4. *Lin, M. L. (2006). Debris flow hazards and mitigations in Taiwan. Keynote Speech, Proc. the 2nd Taiwan-Japan Joint Workshop on Geotechnical Hazards from Large Earthquakes and Heavy Rainfall, Japan.*
5. *Jan, C. D. (2000). Introduction to Debris Flows, Science and Technology Books Company, Taiwan, R. O. C. (in Chinese).*
6. *Huang, L. J. (2001). Introduction to Theory and Practice of Debris-Flow Hazards Mitigation, Chuan-Hwa Publishing Ltd. Taiwan, R.O.C. (in Chinese).*
7. *Yu, F. C., Wang, I. C. and Chen, L.J. (1991). A Study on the Prediction of Potential Debris Flow Hazard Area in the Li-Yu-Tan Watershed. J. Chi. Soil & Water Cons. 22(2), 39-65. (in Chinese).*
8. *Huang, L. J. (2015). Review of the Researches on Debris Flows in Taiwan during Past Thirty Years, International Journal of Emerging Technology and Advanced Engineering, Vol. 4, Issue 12, pp. 122-131.*
9. *Lin, L. K. (2000). Development of Expert Systems for Mitigation of Debris-Flow Hazards. 2nd Int. Conf. on Debris-Flow Hazards Mitigation: Mechanics, Prediction, and Assessment. Taipei, Taiwan, Aug. 16-18. (ed. by Wieczorek, G. G. and Naeser, N. D.) A. A. Balkema, Rotterdam, Brookfield. 425-429.*
10. *Hsiao, C. Y. (2003). The Hazard Analysis of Debris-Flow Zones by Using Neural Network and Digital Terrain Model: Case Study from the Watershed Area of Chen-Yu-Lan River, Nantou, Master Thesis, Institute of Applied Geophisics, National Taiwan Ocean University.*
11. *Chen, K. H. (2006). The Study of Fuzzy Reasoning Applied to the Prediction System of Debris Flows. Master Thesis. Institute of Civil Engineering, Fung-Chia University. (in Chinese).*

12. *Lin, P. S., Lin, J. Y., Lin, S. Y. And Lai, J. R. (2006). Hazard Assessment of Debris Flows by Statistical Analysisand GIS in Central Taiwan, Int. J. Appl. Sci. Engng., 4(2), 165-287.*

13. *Tsai, S. M. (2007). The Study of Case-Based Reasoning Applied to the Prediction System of Debris Flows. Master Thesis. Institute of Civil Engineering, Fung-Chia University. (in Chinese)*

14. *Chen, T. W. (2016). Application of Artificial Neural Networks to Classification, Velocity Distribution and Risk Assessment of Debris-Flows, Master Thesis, Institute of Civil Engineering, National Kaohsiung University of Applied Sciences.*

15. *Lin, Y. R. (2016). Application of Grey System Theory to Risk Assessment of Debris-Flow Hazards, Master Thesis, Institute of Civil Engineering, National Kaohsiung University of Applied Sciences.*

16. *Hagan, M. T., Demuth, H. B. and Beale, M. H. (1996). Neural Network Design, Boston: PWS Publishers, 1996.*

17. *Hecht-Nielsen, R. (1990) Neurocomputing, Addison-Wesley Publishing Company, Inc.*

18. *Jang, J.-S. R., Sun, C.-T. and Mizutani, E.(1997) Neuro-Fuzzy and Soft Computing:A Computational Approach to Learning and Machine Intelligence, Prentice-Hall, Inc.*

19. *Harrington, P. (2012) Machine Learning in Action, Manning Publications.*

20. *Abdi, H. and Williams, L. J. (2010). Principal Component Analysis, Wiley Interdisciplinary Reviews:Computational Statistics, Vol. 2, pp.433-459.*

21. *Shaw, P. J. A. (2003). Multivariate Statistics for the Environmental Sciences, Hodder-Arnold.*